

Anteproyecto: Transcriptor de audio a texto

Juan Esteban Echeverri, Jhonny Steven Ibarbo, Brayan Urdinola Sánchez

Departamento de ingeniería, Universidad Autónoma de Manizales

juane.echeverrip@autonoma.edu.co, jhonnys.ibarboh@autonoma.edu.co,
brayan.urdinolas@autonoma.edu.co

Planteamiento del problema: A pesar de que en la actualidad existen muchas técnicas para analizar y procesar información de una manera óptima, lo cierto es que para lograr esto resulta necesario que la información sea presentada, en la mayoría de ocasiones, en forma de texto. Sin embargo, existen otras formas de mostrar información como por ejemplo, el audio, el cual abarca mucho terreno en el procesamiento de señales, pero poco en el procesamiento de datos, lo que dificulta el análisis de las palabras dicha por una persona, el contenido de una canción, las muletillas de un orador, lo que se dice en una película, serie o video, etc.

¿Cómo se dará solución? El enfoque estará dirigido a usar las herramientas de procesamiento de texto para procesar diferentes audios, esto se hará a través del módulo de Python *Speech Recognition*, mediante el cual se hará la conversión de un archivo de audio a texto para posteriormente realizar el procesamiento mediante Apache Spark o MapReduce y así poder obtener diferente información como por ejemplo: muletillas de un orador, palabras que más se dicen en diferentes géneros de música o incluso la cantidad de palabras en las que suelen rondar las canciones de dicho género. Cabe resaltar que para hacer esto posible, resulta necesaria la importación del archivo de audio, la conversión a formato WAV, la incorporación de ajustes, condiciones y, de ser necesario, filtros para garantizar que la

transcripción sea lo más precisa posible.

Retos:

- Presentar la información de una manera clara, concisa y fácil de entender.
- De ser necesario, la implementación de una interfaz gráfica.
- Garantizar la precisión con la que se reconocen las palabras.
- Lograr el reconocimiento de palabras a pesar de que haya ruido de fondo o alguna melodía.
- Mantener la precisión de la transcripción independientemente del acento o idioma que se reproduzca en el archivo de audio.

Base de datos: Esto dependerá del modo de importación, puesto que si el archivo de audio es importado por el usuario para ser procesado, la base de datos podría ser cualquier sitio de donde se haya extraído dicho archivo, incluso el mismo micrófono, por tanto no estaría definida. Por otro lado, si la importación se realiza mediante un enlace (URL/link), entonces la base de datos será la plataforma de YouTube, puesto que a partir de aquí se procesará el video (suministrado por el enlace) para extraer el audio, convertirlo a formato WAV y realizar todo el procedimiento necesario para aplicarle procesamiento de datos.